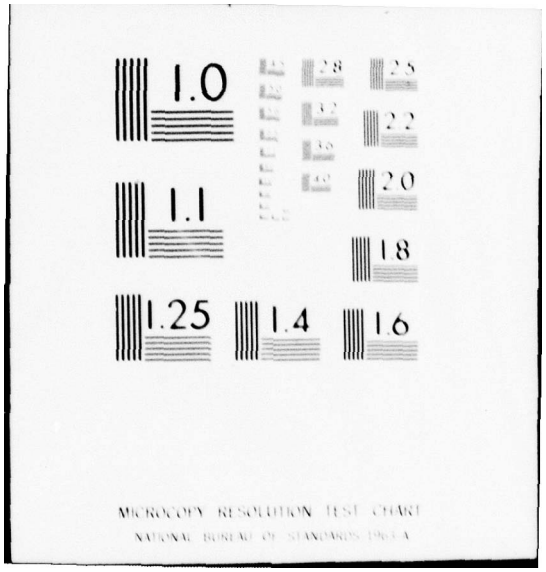
AD-A055 435 PURDUE UNIV LAFAYETTE IND DEPT OF STATISTICS F/G 12/1
CHI-SQUARE TECHNIQUES.(U)
MAY 78 D S MOORE AFOSR-77-3291
UNCLASSIFIED AFOSR-TR-1048 NL

1 OF 1
AD
A055 435

END
DATE FILMED
8-78
DDC


1.0
1.1
1.25
1.4
1.6
2.8
2.5
2.2
2.0
1.8
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 78-1048
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:

FOR FURTHER TRAN

4. TITLE (and Subtitle): CHI-SQUARE TECHNIQUES.
5. TYPE OF REPORT & PERIOD COVERED: Interim rept.
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): David S. Moore
8. CONTRACT OR GRANT NUMBER(s): AFOSR-77-3291
9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Purdue University
Department of Statistics
West Lafayette, IN 47907
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS:
Air Force Office of Scientific Research/NM
Bolling AFB, Washington, DC 20332
12. REPORT DATE: May 1978
13. NUMBER OF PAGES: 37
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): AFOSR 78-1048
15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Goodness of fit
Chi-square tests

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper is a chapter to appear in a handbook on goodness of fit for users of statistics. Both the classical chi-square test and modern variations are discussed. Recommendations are made on the overall usefulness of these tests as well as on specific issues such as choice of cells. Numerical examples are included, using data sets which will appear in the book.

DD FORM 1 JAN 73 1473

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

291 730

(2)

CHAPTER 3

CHI-SQUARE TECHNIQUES

David S. Moore*
Purdue University

1978

3.1 INTRODUCTION

The chi-square test, invented by Karl Pearson in 1900, is not only the oldest test of fit, but the oldest non-trivial test of significance. While it is inferior in power to other classes of tests of fit, the Pearson test is unexcelled in ease and flexibility of use. It applies with little modification to the problems of testing fit to parametric families of distributions, to discrete distributions, and to multivariate distributions. Recent variations of the Pearson statistic have improved the flexibility of chi-square techniques, especially when unknown parameters must be estimated in the hypothesized family. This chapter focuses on those variations of the chi-square which appear most useful to practitioners, with briefer comments and references for other aspects of the subject. Numerical examples are given in Section 3.2.4 for the Pearson statistic and in Section 3.3.3 for some newer chi-square statistics. In addition, Section 3.4.2 illustrates the use of chi-square techniques in the less common situations of multivariate observations and censored data. Recommendations on the use of chi-square techniques in practice appear in Sections 3.2.5 and 3.4.1.

*Preparation of this chapter was supported in part by the Air Force Office of Scientific Research under Grant AFOSR 77-3291. The author is grateful to Dr. Daniel Mihalko for his assistance.

3.2 THE PEARSON CHI-SQUARE STATISTIC

3.2.1 Simple Hypothesis

To test the simple hypothesis that a random sample $X_1,\ldots,X_n$ has the distribution function $F(x)$, Pearson partitioned the range of X_j into M cells, say $E_1,\ldots,E_M$. If $N_1,\ldots,N_M$ are the observed number of X_j's in these cells, then N_i has the binomial distribution with parameters n and

$$p_i = P(X_j \text{ falls in } E_i) = \int_{E_i} dF(x) \tag{3.1}$$

when the null hypothesis is true. Pearson reasoned that the differences $N_i - np_i$ between observed and expected cell frequencies express lack of fit of the data to F, and he sought an appropriate function of these differences for use as a measure of fit.

Pearson's argument here was in three stages: (i) The quantities $N_i - np_i$ have in large samples approximately a multivariate normal distribution, and this distribution is nonsingular if only M-1 of the cells are considered. (ii) If $Y = (Y_1,\ldots,Y_p)'$ has a nonsingular p-variate normal distribution $N_p(\mu,\Sigma)$, then the quadratic form $(Y-\mu)'\Sigma^{-1}(Y-\mu)$ appearing in the exponent of the density function has the $\chi^2(p)$ distribution as a function of Y. Here of course μ is the p-vector of means, and Σ is the pxp covariance matrix of Y. (iii) Computation shows that if $Y = (N_1 - np_1,\ldots,N_{M-1} - np_{M-1})'$, this quadratic form is

$$\chi^2 = \sum_{i=1}^{M} \frac{(N_i - np_i)^2}{np_i},$$

which therefore has approximately the $\chi^2(M-1)$ null distribution in large samples. This is the Pearson chi-square statistic.

This elegant argument will reappear in our survey of recent advances in chi-square tests. Pearson reduced the problem of testing fit to the problem of testing whether a multinomial distribution has cell probabilities p_i given by (3.1). This problem, and the statistic X^2, do not depend on whether F is univariate or multivariate, discrete or continuous. But if F is continuous, consideration of only the cell frequencies N_i does not fully use the information available in the observations X_j. Thus the flexibility and relative lack of power of X^2 stem from the same source.

3.2.2 Composite Hypothesis

It is common to wish to test the composite hypothesis that the distribution function of the observations X_j is a member of a parametric family $\{F(\cdot|\theta): \theta \text{ in } \Omega\}$, where Ω is a p-dimensional parameter space. Pearson recommended estimating θ by an estimator $\tilde{\theta}_n$ (a function of $X_1,\ldots,X_n$), and testing fit to the distribution $F(\cdot|\tilde{\theta}_n)$. Thus the estimated cell probabilities become

$$p_i(\tilde{\theta}_n) = \int_{E_i} dF(x|\tilde{\theta}_n)$$

and the Pearson statistic is

$$X^2(\tilde{\theta}_n) = \sum_{i=1}^{M} \frac{[N_i - np_i(\tilde{\theta}_n)]^2}{np_i(\tilde{\theta}_n)} .$$

Pearson did not think that estimating θ changes the large sample distribution of X^2, at least when $\tilde{\theta}_n$ is consistent. In this he was wrong. It was not until 1924 that Fisher showed that the limiting null distribution of $X^2(\tilde{\theta}_n)$ is not $\chi^2(M-1)$, and that this distribution depends on the method of estimation used.

Fisher argued that the appropriate method of estimation is maximum likelihood estimation based on the cell frequencies N_i. This *grouped data MLE* is the solution of the equations

$$\sum_{i=1}^{M} \frac{N_i}{p_i(\theta)} \frac{\partial p_i(\theta)}{\partial \theta_k} = 0, \qquad k = 1,\ldots,p \tag{3.2}$$

obtained by differentiating the logarithm of the multinomial likelihood function. Fisher showed further that an asymptotically equivalent estimator

can be obtained by choosing θ to minimize $X^2(\theta)$ for the observed N_i. This *minimum chi-square estimator* is the solution of

$$\sum_{i=1}^{M} \left\{\frac{N_i}{p_i(\theta)}\right\}^2 \frac{\partial p_i(\theta)}{\partial \theta_k} = 0, \quad k = 1,\ldots,p. \tag{3.3}$$

Let us denote either estimator by $\bar{\theta}_n$. Then $X^2(\bar{\theta}_n)$ is conceptually the Pearson statistic for testing fit to $F(\cdot|\bar{\theta}_n)$, the member of the family $\{F(x|\theta)\}$ which is closest to the data if the Pearson statistic is used as a measure of distance. Fisher showed that the *Pearson-Fisher statistic* $X^2(\bar{\theta}_n)$ has the $\chi^2(M-p-1)$ distribution under the null hypothesis, no matter what θ in Ω is the true value. This is the famous "lose one degree of freedom for each parameter estimated" result.

Neyman (1949) noted that another estimator asymptotically equivalent to $\bar{\theta}_n$ can be obtained by minimizing the modified chi-square statistic

$$\sum_{i=1}^{M} \frac{[N_i - np_i(\theta)]^2}{N_i}.$$

This *minimum modified chi-square estimator* is the solution of

$$\sum_{i=1}^{M} \frac{p_i(\theta)}{N_i} \frac{\partial p_i(\theta)}{\partial \theta_k} = 0, \qquad k = 1,\ldots,p. \tag{3.4}$$

Since for the purposes of large sample theory this estimator is interchangable with the previous two, call it also $\bar{\theta}_n$ to minimize notation. Neyman's remark is important because equations (3.4) are more often solvable in closed form than are (3.3) and (3.2).

EXAMPLE. Consider the chi-square test of fit to the family of density functions

$$f(x|\theta) = \tfrac{1}{2}(1 + \theta x) \qquad -1 \leq x \leq 1 \tag{3.5}$$

with $\Omega = (-1,1)$. This family has been used as a model for the distribution of the cosine of the scattering angle in some beam-scattering experiments in physics. For cells $E_i = (a_{i-1}, a_i]$ with

$$-1 = a_0 < a_1 < \ldots < a_M = 1,$$

we have

$$p_i(\theta) = \int_{a_{i-1}}^{a_i} f(x|\theta)dx$$

$$= \frac{\theta}{4}(a_i^2 - a_{i-1}^2) + \frac{1}{2}(a_i - a_{i-1}).$$

It is easily seen that neither (3.2) nor (3.3) has a closed solution, while (3.4) has solution

$$\bar{\theta}_n = -2 \frac{\sum_{i=1}^{M} (a_i - a_{i-1})(a_i^2 - a_{i-1}^2)/N_i}{\sum_{i=1}^{M} (a_i^2 - a_{i-1}^2)^2/N_i}.$$

Substituting this value in the Pearson statistic produces an easily computed test of fit for the family (3.5) using $\chi^2(M-2)$ critical points.

But even the minimum modified chi-square estimator must often be obtained by numerical solution of its defining equation. If cells $E_i = (a_{i-1}, a_i]$ are used in a chi-square test of fit to the normal family

$$F(x|\mu,\sigma) = \Phi(\frac{x-\mu}{\sigma}) \qquad -\infty < x < \infty,$$

(Φ is the standard normal distribution function), then

$$p_i(\mu,\sigma) = \Phi(\frac{a_i - \mu}{\sigma}) - \Phi(\frac{a_{i-1} - \mu}{\sigma}).$$

It takes only a moment to see that none of the three versions of $\bar{\theta}_n$ can be obtained algebraically, so that recourse to numerical solution is required. Most computer libraries contain efficient routines using (for example) Newton's method to accomplish the solution.

This circumstance calls to mind Fisher's warning that his "lose one degree of freedom for each parameter estimated" result is <u>not</u> true when estimators not asymptotically the same as $\bar{\theta}_n$ are used. For example, in

testing univariate normality we may not simply use the raw data MLE's

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

$$\hat{\sigma} = \{\frac{1}{n}\sum_{j=1}^{n}(X_j-\bar{X})^2\}^{1/2}$$

in the Pearson statistic. Chernoff and Lehmann (1954) studied the consequences of using the raw data MLE $\hat{\theta}_n$ in the Pearson statistic. They found that $X^2(\hat{\theta}_n)$ has as its limiting distribution under $F(\cdot|\theta)$ the distribution of

$$\chi^2(M-p-1) + \sum_{k=1}^{p} \lambda_k(\theta)\chi_k^2(1). \tag{3.6}$$

Here $\chi^2(M-p-1)$ and $\chi_k^2(1)$ are independent chi-square random variables with the indicated numbers of degrees of freedom. The numbers $\lambda_k(\theta)$ satisfy $0 \leq \lambda_k(\theta) < 1$. So the large sample distribution of $X^2(\hat{\theta}_n)$ is not χ^2 and depends on the true value of θ. All that can be said in general is that the correct critical points fall between those of $\chi^2(M-p-1)$ and those of $\chi^2(M-1)$. These bounds often make $X^2(\hat{\theta}_n)$ usable in practice, especially when the number of cells M is large and the number of parameters p is small.

3.2.3 Choosing Cells In The Pearson Statistic

A major objection to the use of chi-square tests has been the arbitrariness introduced by the necessity to choose cells. This choice is guided by two considerations: the power of the resulting test, and the desire to use the asymptotic distribution of X^2 as an approximation to the exact distribution for sample size n. These issues have been studied in detail for the case of a simple hypothesis, i.e., the case of testing fit to a completely specified distribution F. Recommendations can be made in this case which may reasonably be extended to the case of testing fit to a parametric family $\{F(\cdot|\theta)\}$.

Mann and Wald (1942) initiated the study of the choice of cells in the Pearson test of fit to a continuous distribution F. They recommended, first, that the cells be chosen to have equal probabilities under the hypothesized distribution F. The advantages of such a choice are: (1) The distance $\sup|F_1(x)-F(x)|$ to the nearest alternative indistinguishable from F by X^2 is maximized. (2) The chi-square test is unbiased. (Mann and Wald proved only local unbiasedness, but the test is in fact unbiased against arbitrary alternatives F_1. This is not true when the cells have unequal probabilities under F.) (3) Empirical studies have shown that the χ^2 distribution is a more accurate approximation to the exact null distribution of X^2 when equiprobable cells are employed.

Mann and Wald then made recommendations on the number M of equiprobable cells to be used. Their work rests on large-sample approximations and on a somewhat complex minimax criterion, so that it is at best a rough guide in practice. Mann and Wald found that for a sample of size n (large) and significance level α, one should use approximately

$$M = 4\left\{\frac{2n^2}{c(\alpha)^2}\right\}^{1/5} \tag{3.7}$$

where $c(\alpha)$ is the upper α-point of the standard normal distribution. The optimum is quite broad. In particular, the M of (3.7) can be halved with little effect on power. Retracing the Mann-Wald calculations using better approximations, as in Schorr (1974), confirms that the "optimum" M is smaller than the value given by (3.7). Since the exact optimum depends on the criterion, a choice of error probabilities, and of course on the assumption that the hypothesized F contains no unknown parameters, the practitioner need not go beyond the following recommendation. For $n \geq 50$, choose a number M of equiprobable cells falling between the value (3.7) for $\alpha = 0.05$ and half that value. This recommendation is not an endorsement of the use of $\alpha = 0.05$ in tests

of fit. Because (3.7) increases slowly with α, but overstates the number of cells required, the value for $\alpha = 0.05$ can also be used when larger significance levels are in mind.

For $n < 50$, computation and simulation suggest that the recommendations above remain reasonable, even though their theoretical base in asymptotic theory is no longer valid. For small sample sizes, the question of the accuracy of the χ^2 approximations to the null distribution of the Pearson statistic becomes more prominent and has traditionally influenced the choice of M relative to n. The availability of inexpensive computing power has led to extensive study of this issue since Cochran (1954) gave the commonly accepted rule of thumb. Cochran's rule was that all expected cell frequencies np_i should be at least 1, with at least 80 percent being at least 5. Two papers which summarize more recent work are Roscoe and Byars (1971), a simulation study, and Good, Gover and Mitchell (1970), which is based on computation of the exact distribution. It is notable that current recommendations are stated in terms of the average expected cell frequency rather than in terms of the minimum expected frequency.

Here are the findings of Roscoe and Byars, which may serve as a guide for practitioners.

(i) With equiprobable cells, the average expected cell frequency should be at least 1 (that is, $n \geq M$) when testing fit at the $\alpha = 0.05$ level; for $\alpha = 0.01$, the average expected frequency should be at least 2 (that is, $n \geq 2M$).

(ii) When cells are not approximately equiprobable, the average expected frequencies in (i) should be doubled.

(iii) These recommendations apply when $M \geq 3$. For $M = 2$ (1 degree of freedom), the chi-square test should be replaced by the test based on the exact binomial distribution.

Even guideline (ii) is satisfied whenever Cochran's rule is satisfied, and so is strictly less restrictive. The chi-square test with χ^2 critical points has a true α higher than the nominal α when the guidelines (i) and (ii) are not met. Roscoe and Byars considered only $\alpha = 0.05$ and $\alpha = 0.01$, whereas tests of fit preliminary to other statistical procedures often use $\alpha = 0.25$ or similar levels. Since the χ^2 approximation seems least accurate in the tails, it appears that rule (i) is adequate for such larger values of α. It should be noted, however, that simulation and analytic approximations both suggest that using the maximum number of equiprobable cells ($M = n$) allowed by guideline (i) results in a test with less power than tests having fewer cells, against all but very short-tailed alternatives. Since the Mann-Wald suggestion (3.7) falls within the Roscoe-Byars guidelines, we can reaffirm the recommendations underlined above.

Recommendations in the composite case are less easily made. Both theoretical and empirical results suggest that the choice of M depends on the particular hypothesized family, on the method by which the unknown parameters are estimated, and on the alternatives we wish to detect. The degree of arbitrariness is greatly reduced by using data-dependent cells which are equiprobable under the estimated parameter values. This is possible when the hypothesized family of distributions has only location and scale parameters, as is illustrated by Example 1 in Section 3.2.4 and discussed in Section 3.3.1. The most thorough study to date of the effect of M on the power of a chi-square test in the composite case is Dahiya and Gurland (1973). They investigated the test of univariate normality using $\chi^2(\hat{\theta}_n)$, the Pearson statistic with parameters estimated by the raw data MLE, and data-dependent cells equiprobable under the estimated parameter values

for sample size 50 and 100. Against some alternatives (double exponential, logistic), power decreases as M increases, so that M = 3 is optimal. This surprising result does not hold for the Pearson-Fisher statistic $X^2(\bar{\theta}_n)$, or for the Rao-Robson statistic which we will recommend in Section 3.3.2. For alternatives less close to the normal family, a number of cells roughly half that specified by (3.7) gave the highest power.

The examples in this chapter will use (3.7) for $\alpha = 0.05$ as a guide in choosing M. This avoids subjectivity, and in the author's experience results in greater sensitivity than the calculations of Dahiya and Gurland suggest. There is some evidence that an M half this size may give slightly better power.

3.2.4 Examples Of The Pearson Test

Because of its relative lack of power, X^2 cannot be recommended for testing fit to standard distributions for which special-purpose tests are available, or for which the special tables of critical points needed to apply tests based on the empirical distribution function (EDF) when parameters are estimated have been computed. Testing fit to the family (3.5) is, on the other hand, a realistic application of the Pearson-Fisher statistic $X^2(\bar{\theta}_n)$. Indeed, only chi-square tests allow solution of this problem using tabled critical points. The examples below of X^2 applied to the NOR data set are intended only as illustrations of the mechanics of applying the test.

EXAMPLE 1. Since NOR purports to be data simulating a normal sample with $\mu = 100$ and $\sigma = 10$, let us first assess the simulation by testing fit to this specific distribution. The Mann-Wald recipe (3.7) with $\alpha = 0.05$ and $n = 100$ gives $M = 24$. For computational convenience, we use $M = 25$ cells chosen to be equiprobable under N(100,100). The cell boundaries are $100 + 10z_i$, where z_i is the $0.04i$ point from the standard normal table, $i = 1,2,\ldots,24$. For example, the 0.04 point is -1.75, so the upper boundary of the leftmost cell is $100 + (10)(-1.75) = 82.5$. Table 3.1 shows the cells and their observed frequencies. The expected frequencies are all $(100)(0.04) = 4$. When $p_i = 1/M$ for all i, we have

$$X^2 = \frac{M}{n} \sum_{i=1}^{M} (N_i - \frac{n}{M})^2.$$

So in this example,

$$X^2 = \frac{1}{4} \sum_{i=1}^{25} (N_i - 4)^2$$

$$= \frac{112}{4} = 28.$$

The appropriate distribution is $\chi^2(24)$, and the P-value (attained significance level) of $X^2 = 28$ is 0.260.

To test the NOR data for fit to the family of univariate normal distributions, an intuitively reasonable procedure is to estimate μ,σ by $\bar{X},\hat{\sigma}$ and use cells with boundaries $\bar{X} + z_i\hat{\sigma}$, where z_i are as before. These cells are equiprobable under the normal distribution with $\mu = \bar{X}$ and $\sigma = \hat{\sigma}$. It will be remarked in Section 3.3 that the Pearson statistic with these data-dependent cells has the same large sample distribution as if the fixed cell boundaries $100 + 10z_k$ to which the random boundaries coverage were used. This distribution is <u>not</u> $\chi^2(24)$, since μ and σ were estimated by their raw data MLE's $\bar{X}$ and $\hat{\sigma}$ in computing the cell probabilities $p_i(\bar{X},\hat{\sigma}) = 0.04$. The appropriate distribution has the form (3.6), so that its critical points fall between those of $\chi^2(24)$ and $\chi^2(22)$. Calculation shows that $\bar{X} = 99.54$ and $\hat{\sigma} = 10.46$. The cell boundaries $\bar{X} + \hat{\sigma}z_k$ and the observed cell frequencies are given at the right of Table 3.1. The observed chi-square value is $X^2 = 22$, reflecting the somewhat better fit when parameters are estimated from the data. The P-value falls between 0.460 (from $\chi^2(22)$) and 0.579 (from $\chi^2(24)$).

For comparison, the same procedure was applied to test the LOG data set for normality. In this case, $\bar{X} = 99.84$ and $\hat{\sigma} = 16.51$, and the observed chi-square value using cell boundaries $\bar{X} + \hat{\sigma}z_k$ is $X^2 = 31.5$. The corresponding

TABLE 3.1

Chi-square tests for

normality of the NOR data

	Fit to N(100,100)		Fit to normal family	
Cell	Upper Boundary	Frequency	Upper Boundary	Frequency
1	82.5	3	81.2	3
2	85.9	8	84.8	5
3	88.3	5	87.3	5
4	90.1	8	89.2	5
5	91.6	4	90.7	6
6	92.9	2	92.1	4
7	94.2	1	93.5	3
8	95.3	5	94.6	1
9	96.4	6	95.8	4
10	97.5	1	96.9	6
11	98.5	3	98.0	3
12	99.5	3	99.0	3
13	100.5	4	100.1	2
14	101.5	2	101.1	5
15	102.5	2	102.2	2
16	103.6	7	103.3	5
17	104.7	7	104.5	9
18	105.8	3	105.6	3
19	107.1	1	107.0	1
20	108.4	2	108.3	1
21	109.9	4	109.9	5
22	111.7	6	111.8	6
23	114.1	6	114.3	6
24	117.5	4	117.8	4
25	∞	3	∞	3

P-value lies between 0.086 (from $\chi^2(22)$) and 0.140 (from $\chi^2(24)$). Thus this test has correctly concluded that NOR fits the normal family well, while the fit of LOG is marginal. Since the logistic distributions are difficult to distinguish from the normal family, this is a pleasing performance. In contrast, the same procedure with M = 10 has $X^2 = 9.4$ for the LOG data, so that the P-value lies between 0.225 (from $\chi^2(7)$) and 0.402 (from $\chi^2(9)$). Using 3 cells gives $X^2 = 0.98$ and again fails to suggest that the LOG data set is not normally distributed. Thus for these particular data, the larger M suggested by (3.7) produces a more sensitive test.

EXAMPLE 2. The same procedure can be applied to the EMEA data, but a glance shows that these data as given are discrete and therefore not normal. Indeed, with 15 cells equiprobable under the $N(\bar{X},\hat{\sigma})$ distribution for these data, $X^2 = 554$. Since the data are grouped in classes centered at integers, a more intelligent procedure is to use fixed cells of unit width centered at the integers, with cell probabilities computed from $N(\bar{X},\hat{\sigma})$. Of course, $\bar{X}$ and $\hat{\sigma}$ from the grouped data are only approximate. Sheppard's correction for $\hat{\sigma}$ improves the approximation, and gives $\bar{X} = 14.540$ and $\hat{\sigma} = 2.216$. Calculating the cell probabilities and computing the Pearson statistic, we obtain $X^2 = 7.56$. The P-value lies between 0.819 (from $\chi^2(12)$) and 0.911 (from $\chi^2(14)$), so that the EMEA data fit the normal family very well indeed. The applicability of X^2 to grouped data such as these is an advantage of chi-square methods.

3.2.5 Recommendations For Use Of The Pearson Statistic

(1) When the raw data MLE $\hat{\theta}_n$ is computationally simpler than the grouped data estimator $\bar{\theta}_n$, do not hesitate to use $\hat{\theta}_n$. The critical points of $X^2(\hat{\theta}_n)$ fall between those of $\chi^2(M-1)$ and $\chi^2(M-p-1)$, and

simulation suggests that $X^2(\hat{\theta}_n)$ is usually more powerful than the Pearson-Fisher test based on $X^2(\bar{\theta}_n)$ and $\chi^2(M-p-1)$ critical points.

(2) When testing fit to a location-scale family $\{F(\cdot|\theta)\}$, use cells which are equiprobable under the estimated value of θ. The fact that these cells are data-dependent does not affect the distribution theory, as Section 3.3.1 discusses more fully.

(3) Choose the number M of equiprobable cells to be approximately $2n^{2/5}$. (This is based on the discussion in Section 3.2.3. Half the Mann-Wald recipe (3.7) for $\alpha = 0.05$ is $1.9n^{2/5}$.)

3.3 GENERAL CHI-SQUARE STATISTICS

3.3.1 Data-dependent Cells

As already noted in Section 3.2.4, the use of data-dependent cells increases the flexibility of chi-square tests, fortunately without increasing their complexity in practice. The essential requirement is that as the sample size increases, the random cell boundaries must converge in probability to a set of fixed boundaries. The limiting cells will usually be unknown, since they depend on the true parameter value θ_0. Random cells are used in chi-square tests by "forgetting" that the cells are data-dependent and proceeding as if fixed cells had been chosen. Since the cell frequencies are no longer multinomial, the theory of such tests is mathematically difficult. But in practice, the limiting distribution of X^2 with random cells is exactly the same as if the limiting fixed cells had been used. This is true even when parameters are estimated. Details and regularity conditions appear in Section 4 of Moore and Spruill (1975). Therefore, any statistic, such as the Pearson-Fisher $X^2(\bar{\theta}_n)$, which has a θ_0-free limiting null distribution using fixed cells, has that same limiting null distribution for any choice of converging random cells.

A statistic such as the Chernoff-Lehmann $X^2(\hat{\theta}_n)$ which has a θ_0-dependent limiting null distribution for fixed cells, has in general this same

deficiency with random cells. But if the hypothesized family $\{F(\cdot|\theta)\}$ is a location-scale family, a proper choice of random cells eliminates this θ_0-dependency and also allows cells to be chosen equiprobable under the estimated θ, thus matching the recommended practice in the simple hypothesis case. Such cell choices should be made whenever possible. Theorem 4.3 of Moore and Spruill (1975) is a general account of this. Let us here illustrate it by returning to the X^2 statistic for testing univariate normality.

When the parameter $\theta = (\mu,\sigma)$ is estimated by $\hat{\theta}_n = (\bar{X},\hat{\sigma})$ and cell boundaries $\bar{X} + z_i\hat{\sigma}$ are used, the estimated cell probabilities are

$$p_i(\bar{X},\hat{\sigma}) = \int_{\bar{X} + z_{i-1}\hat{\sigma}}^{\bar{X} + z_i\hat{\sigma}} (2\pi\hat{\sigma}^2)^{-1/2} e^{-(t-\bar{X})^2/2\hat{\sigma}^2} dt$$

$$= \int_{z_{i-1}}^{z_i} (2\pi)^{-1/2} e^{-u^2/2}\, du$$

These are not dependent on $(\bar{X},\hat{\sigma})$, and are equiprobable if z_i are the successive i/M points of the standard normal distribution. Since this choice of cells leaves both N_i and p_i unchanged when any location-scale transformation is applied to all observations X_j, the Pearson statistic has the same distribution for all (μ,σ). The limiting null distribution has the form (3.6) but the λ_k are now free of any unknown parameter. Critical points may therefore be computed. Two methods for doing so, and tables for testing normality, appear in Dahiya and Gurland (1972) and Moore (1971). Dahiya and Gurland (1973) study the power of this test. The idea of using random cells in this fashion is due to A. R. Roy (1956) and G. S. Watson (1957, 1958, 1959). We will refer to the Pearson statistic using the raw data MLE and random cells as the *Watson-Roy statistic*. Example 1 in Section 3.2.4 illustrated its use.

Note that the Watson-Roy statistic has θ-free limiting null distribution only for location-scale families, that this distribution is not a standard tabled distribution, and that a separate calculation of critical points is required for testing fit to each location-scale family. These statements are also true for EDF tests of fit. Since the latter are more powerful, the Watson-Roy statistic has few advantages when $F(\cdot|\theta)$ is univariate and continuous. Nonetheless, data-dependent cells move the cells to the data without essentially changing the asymptotic distribution theory of the chi-square statistic. They should be routinely employed in practice, and this is done in most of the examples in this chapter.

3.3.2 General Quadratic Forms

Some of the most useful recent work on chi-square tests involves the study of quadratic forms in the standardized cell frequencies other than the sum of squares used by Pearson. Random cells are commonly recommended in these statistics, for the reasons outlined in Section 3.3.1, and do not affect the theory. A statement of the nature and behavior of these general statistics of chi-square type is necessarily somewhat complex. Practitioners may find it helpful to study the examples computed in Section 3.3.3 and in Rao and Robson (1974) before approaching the summary treatment below.

Random cells should be denoted by $E_{in}(X_1,\ldots,X_n)$ in a precise notation, but here the notation E_i for cells and N_i for cell frequencies will be continued. The "cell probabilities" under $F(\cdot|\theta)$ are

$$p_i(\theta) = \int_{E_i} dF(x|\theta) \qquad i = 1,\ldots,M.$$

Denote by $V_n(\theta)$ the M-vector of standardized cell frequencies having ith component

$$[N_i - np_i(\theta)]/(np_i(\theta))^{1/2}.$$

If $Q_n = Q_n(X_1,\ldots,X_n)$ is a possibly data-dependent MxM symmetric nonnegative definite matrix, the general form of statistic to be considered is

$$V_n(\tilde{\theta}_n)'Q_nV_n(\tilde{\theta}_n) \tag{3.8}$$

when θ is estimated by $\tilde{\theta}_n$. The Pearson statistic is the special case for which $Q_n \equiv I_M$, the MxM identity matrix. The large-sample theory of these statistics is given in Moore and Spruill (1975). The basic idea is that of Pearson's proof: To show that $V_n(\tilde{\theta}_n)$ is asymptotically multivariate normal (even with random cells) and then apply the distribution theory of quadratic forms in multivariate normal random variables. All statistics of form (3.8) have as their limiting null distribution that of a linear combination of independent chi-square random variables. References on the calculation of such distributions may be found in Davis (1977).

To avoid the necessity to compute special critical points, it is advantageous to seek statistics (3.8) which have a chi-square limiting null distribution. This idea is due to D. S. Robson. Rao and Robson (1974) treat the important case of raw data MLE's. They give the quadratic form in $V_n(\hat{\theta}_n)$ having the $\chi^2(M-1)$ limiting null distribution. The appropriate matrix is $Q(\hat{\theta}_n)$, where

$$Q(\theta) = I_M + B(\theta)[J(\theta)-B(\theta)'B(\theta)]^{-1}B(\theta)',$$

$J(\theta)$ is the pxp Fisher information matrix for $F(\cdot|\theta)$, and $B(\theta)$ is the Mxp matrix with (i,j)th entry

$$p_i(\theta)^{-1/2}\frac{\partial p_i(\theta)}{\partial \theta_j}.$$

The *Rao-Robson statistic* is

$$R_n = V_n(\hat{\theta}_n)'Q(\hat{\theta}_n)V_n(\hat{\theta}_n).$$

This test can be used whenever J-B'B is positive definite. Since nJ is the information matrix from the raw data and nB'B the information matrix from the cell frequencies, J-B'B is always nonnegative definite. Notice that R_n is just the Pearson statistic $X^2(\hat{\theta}_n)$ plus a term which

conceptually builds up the distribution (3.6) to $\chi^2(M-1)$. This term simplifies considerably, since $\sum_1^M \partial p_i/\partial\theta_j = 0$ implies that

$$V_n'B = n^{-1/2}\left(\sum_{i=1}^{M} \frac{N_i}{p_i}\frac{\partial p_i}{\partial\theta_1}, \ldots, \sum_{i=1}^{M} \frac{N_i}{p_i}\frac{\partial p_i}{\partial\theta_p}\right) \tag{3.9}$$

and

$$R_n = X^2(\hat{\theta}_n) + (V'_nB)(J-B'B)^{-1}(V'_nB)', \tag{3.10}$$

all terms being evaluated at $\theta = \hat{\theta}_n$. Further simplification can be achieved in location-scale cases by the use of random cells for which $p_i(\hat{\theta}_n) = 1/M$. Rao and Robson (1974) give several examples of the use of this statistic, using random cells in some cases.

Simulations by Rao and Robson show that R_n has generally greater power than either the Pearson-Fisher or Watson-Roy statistics. Spruill (1975) gives a theoretical treatment showing that R_n dominates the Watson-Roy statistic for any location-scale family $\{F(\cdot|\theta)\}$. Since R_n is powerful, has tabled critical points, and is easy to compute whenever the MLE $\hat{\theta}_n$ can be obtained, it is recommended as a standard chi-square test of fit. Moore (1977) gives a general recipe for the quadratic form having the $\chi^2(M-1)$ distribution when nearly arbitrary estimators $\tilde{\theta}_n$ are used. The idea parallels Pearson's proof, using a generalized inverse of the covariance matrix. The Pearson-Fisher and Rao-Robson statistics are the $\bar{\theta}_n$ and $\hat{\theta}_n$ special cases of this recipe, which is the *Wald's method statistic*.

If (3.6) can be built up to $\chi^2(M-1)$, it can also be chopped down to $\chi^2(M-p-1)$. Dzhaparidze and Nikulin (1974) point out that the appropriate statistic is

$$Z_n(\tilde{\theta}_n) = V'_n(I_M - B(B'B)^{-1}B')V_n$$

where V_n and B_n are evaluated at $\theta = \tilde{\theta}_n$. Z_n has the $\chi^2(M-p-1)$ limiting distribution whenever $\tilde{\theta}_n$ approaches θ_0 at the usual $n^{1/2}$ rate, and can therefore be used with any reasonable estimator of θ. Computation of Z_n is again simplified by (3.9). As might be expected, simulations suggest

that $Z_n(\tilde{\theta}_n)$ is inferior in power to both the Watson-Roy and Rao-Robson statistics.

3.3.3 Examples Of General Chi-Square Tests

EXAMPLE 1. It is desired to test fit to the negative exponential family

$$f(x|\theta) = \theta^{-1}e^{-x/\theta}, \quad 0 < x < \infty$$

where $\Omega = \{\theta: 0 < \theta < \infty\}$. Since the MLE of θ, $\hat{\theta}_n = \bar{X}$, is available, the Rao-Robson statistic is the recommended chi-square test. When p = 1, (3.9) and (3.10) reduce to

$$R_n = \sum_{i=1}^{M} \frac{(N_i - np_i)^2}{np_i} + \frac{1}{nD}\left(\sum_{i=1}^{M} \frac{N_i}{p_i}\frac{dp_i}{d\theta}\right)^2$$

where

$$D = J - \sum_{i=1}^{M} \frac{1}{p_i}\left(\frac{dp_i}{d\theta}\right)^2$$

and J, p_i, $dp_i/d\theta$ are all evaluated at $\theta = \hat{\theta}_n$. For a sample of size n = 100, we will once more use M = 25 equiprobable cells. In this scale-parameter family, equiprobable cells are achieved by the use of random cell boundaries of the form $z_i\bar{X}$. From

$$p_i(\theta) = \int_{z_{i-1}\bar{X}}^{z_i\bar{X}} \theta^{-1}e^{-x/\theta}dx \qquad (3.11)$$

the condition $p_i(\bar{X}) \equiv 1/25$ gives $z_0 = 0$, $z_{25} = \infty$ and

$$z_i = -\log\left(1-\frac{i}{25}\right) \qquad i = 1,\ldots,24.$$

Differentiating (3.11) under the integral sign, then subsituting $\theta = \bar{X}$, gives

$$\frac{dp_i}{d\theta} = \bar{X}^{-1}\left[\left(1-\frac{i}{25}\right)\log\left(1-\frac{i}{25}\right) - \left(1-\frac{i-1}{25}\right)\log\left(1-\frac{i-1}{25}\right)\right] = v_i/\bar{X}.$$

Because of their iterative nature, the quantities, v_i are easily computed on a programmable calculator. The Fisher information is $J(\theta) = \theta^{-2}$ so that

TABLE 3.2

The Rao-Robson test for the negative exponential family, with 25 equiprobable cells

			WE2		EXP	
i	z_i	v_i	$z_i\bar{X}$	N_i	$z_i\bar{X}$	N_i
1	.0408	-.0392	0.036	1	0.221	6
2	.0834	-.0375	0.073	0	0.451	5
3	.1278	-.0358	0.112	1	0.692	3
4	.1743	-.0340	0.153	1	0.944	2
5	.2231	-.0321	0.196	3	1.208	5
6	.2744	-.0301	0.241	1	1.486	5
7	.3285	-.0279	0.288	2	1.779	7
8	.3857	-.0257	0.338	3	2.088	2
9	.4463	-.0234	0.392	5	2.416	4
10	.5108	-.0209	0.448	5	2.766	3
11	.5798	-.0182	0.509	1	3.140	3
12	.6539	-.0153	0.574	5	3.541	4
13	.7340	-.0123	0.644	3	3.974	6
14	.8210	-.0089	0.721	5	4.445	3
15	.9163	-.0053	0.804	8	4.962	4
16	1.0216	-.0013	0.897	4	5.532	4
17	1.1394	.0032	1.000	16	6.170	3
18	1.2730	.0082	1.118	9	6.893	3
19	1.4271	.0139	1.253	11	7.728	4
20	1.6094	.0206	1.413	7	8.715	2
21	1.8326	.0287	1.609	5	9.923	7
22	2.1203	.0388	1.861	1	11.481	3
23	2.5257	.0524	2.217	3	13.676	3
24	3.2189	.0733	2.826	0	17.430	6
25	∞	.1288	∞	0	∞	3

$$D = \bar{X}^{-2}[1-25\sum_{i=1}^{25} v_i^2]$$

Finally

$$R_{100} = \frac{1}{4}\sum_{i=1}^{25}(N_i-4)^2 + \frac{(25)^2}{100}\,\frac{(\sum_1^{25} N_i v_i)^2}{1-25\sum_1^{25} v_i^2}\,.$$

Table 3.2 records z_i and v_i, from which

$$1-25\textstyle\sum_1^{25} v_i^2 = 0.04255.$$

For the WE2 data set, $\bar{X} = 0.878$. The resulting cell boundaries and cell frequencies appear in Table 3.2, and

$$R_{100} = \frac{1}{4}(351) + \frac{(25)^2}{100}\,\frac{(-0.0519)^2}{0.04255}$$

$$= 87.75 + 0.40 = 88.15$$

This gives a P-value of 3×10^{-9} using the $\chi^2(24)$ distribution. In contrast, the EXP data set has $\bar{X} = 5.415$, cell boundaries and frequencies given at the right of Table 3.2, and

$$R_{100} = \frac{1}{4}(54) + \frac{(25)^2}{100}\,\frac{(-0.1231)^2}{0.04255}$$

$$= 13.5 + 2.23 = 15.73.$$

The P-value from $\chi^2(24)$ is 0.898.

As these examples suggest, the Pearson statistic $X^2(\hat{\theta}_n)$, which is the first component of R_n, is usually adequate for drawing conclusions when M is large and p is small. In this example, the critical points of $X^2(\hat{\theta}_n)$ fall between those of $\chi^2(22)$ and those of $\chi^2(24)$. A reasonable strategy is to compute $X^2(\hat{\theta}_n)$ first, completing the computation of R_n only if the results after the first stage are ambiguous.

EXAMPLE 2. The BAEN data are to be tested for fit to the double-exponential family

$$f(x|\theta) = \frac{1}{2\theta_2}\, e^{-|x-\theta_1|/\theta_2} \qquad -\infty < x < \infty$$

$$\Omega = \{(\theta_1,\theta_2): -\infty < \theta_1 < \infty,\ 0 < \theta_2 < \infty\}.$$

The MLE $\hat{\theta}_n = (\hat{\theta}_{1n}, \hat{\theta}_{2n})$ from a random sample $X_1, \ldots, X_n$ is

$$\hat{\theta}_{1n} = \text{median } (X_1, \ldots, X_n)$$

$$\hat{\theta}_{2n} = \frac{1}{n} \sum_{j=1}^{n} |X_j - \hat{\theta}_{1n}|.$$

In this location-scale setting, equiprobable cells with boundaries $\hat{\theta}_{1n} + a_i \hat{\theta}_{2n}$ will again be employed. Using an even number of cells, say $M = 2\nu$, and choosing the a_i symmetrically as $a_{\nu+i} = -a_{\nu-i} = c_i$, where

$$c_i = -\log(1 - \frac{i}{\nu}) \qquad i = 0, \ldots, \nu$$

(in particular, $a_0 = -\infty$, $a_\nu = 0$, $a_M = \infty$) gives $p_i(\hat{\theta}_n) \equiv 1/M$.

Computations similar to those shown in Example 1 yield

$$\frac{\partial p_i}{\partial \theta_1}(\hat{\theta}_n) = -1/M\hat{\theta}_{2n} \qquad i = 1, \ldots, \nu$$

$$= 1/M\hat{\theta}_{2n} \qquad i = \nu+1, \ldots, M \qquad (3.12)$$

$$\frac{\partial p_i}{\partial \theta_2}(\hat{\theta}_n) = \frac{1}{2\hat{\theta}_{2n}}(c_{k-1}e^{-c_{k-1}} - c_k e^{-c_k}) \qquad i = \nu+k, \ \nu-k+1 \qquad k = 1, \ldots, \nu$$

If $d_k = c_{k-1}e^{-c_{k-1}} - c_k e^{-c_k}$, then

$$B(\hat{\theta}_n)'B(\hat{\theta}_n) = \hat{\theta}_{2n}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & \nu \sum_1^\nu d_i^2 \end{pmatrix}.$$

Since the information matrix is $\theta_2^{-1} I_2$, the matrix $J(\hat{\theta}_n) - B(\hat{\theta}_n)'B(\hat{\theta}_n)$ has rank 1 and the Rao-Robson statistic is not defined. (The reason for this unusual situation is that for this choice of cells, the median is both the

raw data MLE and the grouped data MLE for θ_1.) The Dzhaparidze-Nikulin statistic is

$$Z_n(\hat{\theta}_n) = \frac{M}{n}\sum_{i=1}^{M}(N_i - \frac{n}{M})^2 - \frac{M}{n}\frac{1}{2\sum_1^{\nu} d_i^2}[\sum_{i=1}^{\nu} d_i(N_{\nu+i} + N_{\nu-i+1})]^2$$

This computation was simplified by the fact that B'B is diagonal and the first term of (3.9) is 0 by (3.12) and the definition of the median.

The BAEN data contain n = 33 observations, for which $\hat{\theta}_{1n} = 10.13$ and $\hat{\theta}_{2n} = 3.36$. Table 3.3 contains c_i, upper cell boundaries $\hat{\theta}_{1n} + c_i\hat{\theta}_{2n}$, and cell frequencies for these data. The statistic Z_n is, after some arithmetic,

$$Z_n = \frac{10}{33}\sum_{i=1}^{10}(N_i - 3.3)^2 - \frac{10}{33}\frac{1}{(2)(.1574)}[-1.2828]^2$$

$$= 7.30 - 1.59 = 5.71$$

The P-value from $\chi^2(7)$ is 0.426. The Pearson

TABLE 3.3

Testing the fit of the BAEN data to the double exponential family

Cell	c_i	$\hat{\theta}_{1n} + c_i\hat{\theta}_{2n}$	N_i
1	-1.609	4.722	4
2	-0.916	7.051	7
3	-0.511	8.414	3
4	-0.223	9.380	2
5	0	10.130	1
6	0.223	10.880	3
7	0.511	11.846	4
8	0.916	13.209	3
9	1.609	15.538	4
10	∞	∞	2

statistic $X^2 = 7.30$ has critical points falling between those of $\chi^2(7)$ and

$\chi^2(8)$, taking advantage of the fact that the grouped data MLE was used to estimate one of the two unknown parameters. The corresponding bounds on the P-value are 0.398 and 0.505. The double exponential model clearly fits the BAEN data very well. Even though an anomaly reduced from 2 to 1 the difference in the degrees of freedom of the χ^2 distributions bounding X^2, there is a considerable spread in the corresponding P-values. This is typical when n (and therefore M) is small. In examples where the goodness of fit is less clear than here, use of R_n or Z_n can be essential to a clear conclusion.

3.3.4 Nonstandard Chi-Square Statistics

The class of *standard chi-square statistics* is composed of all nonnegative definite quadratic forms in the standardized cell frequencies, with possibly estimated parameters and data-dependent cells. Such statistics have a unified large-sample theory given by Moore and Spruill (1975). Other classes of statistics are less well explored but may hold promise. A few are mentioned here. Since none can yet compete with standard statistics in practice, this section can be considered optional reading.

(a) Increasing M with n. Standard statistics assume that the number of cells M remains fixed as the sample size n increases. The usual practice is to use more cells as n increases (recall the Mann-Wald suggestion (3.7)), yet this practice is not explicitly recognized in the theory of standard chi-square statistics. Kempthorne (1968) proposed the use of the Pearson statistic with M = n equiprobable cells. Such statistics have a large-sample theory very different from that of standard statistics. For the case of testing fit to a completely

specified distribution, Morris (1975) shows that the Pearson statistic has a normal limiting null distribution in some generality when M increases with n. The behavior of such statistics when parameters must be estimated is largely unexplored. Simulation studies of Kempthorne's statistic suggest that standard statistics with fewer cells have superior power except against very short-tailed alternatives.

(b) Sequentially adjusted cells. By use of the conditional probability integral transformation (see Chapter 6), O'Reilly and Quesenberry (1973) obtain particular members of the following class of nonstandard chi-square tests. Rather than base cell frequencies on cells E_i (fixed) or E_{in} $(X_1,\ldots,X_n)$ (data-dependent) into which all of $X_1,\ldots,X_n$ are classified, the cells used to classify each successive X_j are functions E_{ij} of $X_1,\ldots,X_j$ only. Thus additional observations do not require reclassification of earlier observations, as in the usual random cell case. No general theory of chi-square statistics based on such sequentially adjusted cells is known. O'Reilly and Quesenberry obtain by their transformation approach specific functions E_{ij} such that the cell frequencies are multinomially distributed and the Pearson statistic has the $\chi^2(M-1)$ limiting null distribution. The transformation approach requires the computation of the minimum variance unbiased estimator of $F(\cdot|\theta)$. Testing fit to an uncommon family thus requires the practitioner to do a hard calculation. Moreover, any test using sequentially adjusted cells has the disadvantage that the value of the statistic depends on the order in which the observations were obtained. These are serious barriers to use.

(c) Easterling's approach. Easterling (1976) provides an interesting approach to parameter estimation based on tests of fit. Roughly speaking, he advocates replacing the usual confidence intervals for θ in $F(\cdot|\theta)$ based on the acceptance regions of a test of

$$H_0: \theta = \theta_0$$
$$H_1: \theta \neq \theta_0$$

with intervals based on the acceptance regions of tests of fit to completely specified distributions,

$$H_0^*: G(\cdot) = F(\cdot|\theta_0)$$
$$H_1^*: G(\cdot) \neq F(\cdot|\theta_0) \ .$$

In the course of his discussion, Easterling suggests rejecting the family $\{F(\mathbf{x}|\theta): \theta \text{ in } \Omega\}$ as a model for the data if the (say) 50% confidence interval for θ based on acceptance regions for H_0^* is empty. This "implicit test of fit" deserves comment, using the chi-square case to make some observations which apply as well when other tests of H_0^* are employed.

Taking then the standard chi-square statistic for H_0^*,

$$X^2(\theta_0) = \sum_{i=1}^{M} \frac{[N_{ni} - np_i(\theta_0)]^2}{np_i(\theta_0)} \ ,$$

and denoting by $\chi_\alpha^2(M-1)$ the upper α-point of the $\chi^2(M-1)$ distribution, the $(1-\alpha)$-confidence interval is empty if and only if

$$X^2(\theta) > \chi_\alpha^2(M-1) \qquad \text{for all } \theta \text{ in } \Omega. \tag{3.13}$$

But if $\bar{\theta}_n$ is the minimum chi-square estimator, (3.13) holds if and only if

$$X^2(\bar{\theta}_n) > \chi_\alpha^2(M-1) \ . \tag{3.14}$$

When any $F(x|\theta)$ is true, $X^2(\bar{\theta}_n)$ has the $\chi^2(M-m-1)$ distribution, and the probability of the event (3.14) can be explicitly computed. It is less than α, but close to α when M is large. Thus Easterling's suggestion essentially reduces to the use of standard tests of fit with parameters estimated by the minimum distance method corresponding to the test statistic employed. Moreover, his method by-passes a proper consideration of the distributional effects of estimating unknown parameters.

3.4 RECOMMENDATIONS AND FURTHER EXAMPLES

3.4.1 The Use of Chi-Square Tests

Chi-square tests are generally less powerful than EDF tests and special-purpose tests of fit. It is difficult to assess the seriousness of this lack of power from published sources. Comparative studies have generally used the Pearson statistic rather than the more powerful Watson-Roy and Rao-Robson statistics. Moreover, such studies have often dealt with problems of parameter estimation in ways which tend to understate the power of general purpose tests such as chi-square and Kolmogorov-Smirnov tests. This is true of the study by Shapiro, Wilk and Chen (1968), for example. Reliable information about the power of chi-square tests for normality can be gained from Table IV of Rao and Robson (1974) and from Tables 1 and 2 of Dahiya and Gurland (1973). The former demonstrates strikingly the gain in power (always at least 40% in the cases considered, and usually much greater) obtained by abondoning the Pearson-Fisher statistic for more modern chi-square statistics. Nonetheless, chi-square tests cannot in general match EDF and special

purpose tests of fit in power.

This relative lack of power implies three theses on the practical use of chi-square techniques. First, chi-square tests of fit must compete for use primarily on the basis of flexibility and ease of use. Discrete and/or multivariate data do not discomfit chi-square methods, and the necessity to estimate unknown parameters is more easily dealt with by chi-square tests than by other tests of fit.

Second, chi-square statistics actually having a (limiting) chi-square null distribution have a much stronger claim to practical usefulness. Ease of use requires the ability to obtain (1) the observed value of the test statistic, and (2) critical points for the test statistic. The calculations required for (1) in chi-square statistics are at most iterative solutions of nonlinear equations and evaluation of quadratic forms, perhaps with matrix expressed as the inverse of a given symmetric pd matrix. These are not serious barriers to practical use, given the current availability of computer library routines. Computation of critical points of an untabled distribution is a much harder task for a user of statistical methods. Chi-square and EDF statistics both have as their limiting null distributions the distributions of linear combinations of central chi-square random variables. General statistics of both classes require a separate table of critical points for each hypothesized family. The effort needed is justified when the hypothesized family is common, but should be expended on a test more powerful than chi-square tests. In less common

cases, or when no more powerful test with θ-free null distribution is available, there are several chi-square tests requiring only tables of the χ^2 distribution. These include the Pearson-Fisher, Rao-Robson, and Dzhaparidze-Nikulin tests, and others which can be constructed by the method of Moore (1977). Among the chi-square statistics proposed and studied to date, the Rao-Robson statistic R_n of (3.10) appears to have generally superior power and is therefore the statistic of choice. Computation of R_n in the nonstandard cases most appropriate for chi-square tests of fit does require some mathematical work. However, the Pearson statistic $X^2(\hat{\theta}_n)$ with raw-data MLE's is the first and usually dominant component of R_n. If $X^2(\hat{\theta}_n)$ itself lies in the upper tail of the $\chi^2(M-1)$ distribution, the fit can be rejected without computing R_n.

The third thesis rests on the exposition and examples in this chapter. Chi-square tests are the most practical tests of fit in many situations. When parameters must be estimated in non-location-scale families or in uncommon distributions, when the data are discrete, multivariate, or even (see the next section) censored, chi-square tests remain easily applicable.

3.4.2 Further Examples

Chi-square tests should not be used for testing the fit of full ungrouped samples to common univariate distributions. There are more powerful tests available in such situations. Yet many of the examples given have concerned such situations. This section illustrates the flexibility of chi-square methods in two more appealing settings, one multivariate and one with censored data. As in the earlier examples of this chapter, the required numerical calculations are easily done on a programmable calculator.

EXAMPLE 1. The circular bivariate normal family is a common model for errors in "bombing" a target. It represents the effect of independent normal horizontal and vertical components with equal variances. The density function is

$$f(x,y|\theta) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}\{(x-\mu_1)^2 + (y-\mu_2)^2\}} \qquad -\infty < x,y < \infty$$

$$\Omega = \{\theta = (\mu_1,\mu_2,\sigma): -\infty < \mu_1,\mu_2 < \infty, 0 < \sigma < \infty\}.$$

The MLE of θ from a random sample $(X_1,Y_1),\ldots,(X_n,Y_n)$ is $\hat{\theta}_n = (\hat{\mu}_1,\hat{\mu}_2,\hat{\sigma})$, where

$$\hat{\mu}_1 = \bar{X} \qquad \hat{\mu}_2 = \bar{Y}$$

$$\hat{\sigma}^2 = \frac{1}{2n}\{\sum_{j=1}^{n}(X_j-\bar{X})^2 + \sum_{j=1}^{n}(Y_j-\bar{Y})^2\}.$$

In constructing a test of fit to this family, it is natural to use as cells annuli centered at $(\bar{X},\bar{Y})$ with successive radii $c_i\hat{\sigma}$ for

$$0 = c_0 < c_1 < \ldots < c_{M-1} < c_M = \infty.$$

Thus

$$E_i = \{(x,y): c_{i-1}^2\hat{\sigma}^2 \leq (x-\bar{X})^2 + (y-\bar{Y})^2 < c_i^2\hat{\sigma}^2\}.$$

The cell probabilities are

$$p_i(\theta) = \iint_{E_i} f(x,y|\theta)dxdy$$

and calculation shows that $p_i(\hat{\theta}_n) \equiv 1/M$ when

$$c_i = \{-2\log(1-\frac{i}{M})\}^{1/2} \qquad i = 1,\ldots,M-1.$$

The recommended test is based on the Rao-Robson statistic. Differentiating $p_i(\theta)$ under the integral sign, then substituting $\theta = \hat{\theta}_n$ gives

$$\frac{\partial p_i}{\partial \mu_1}\Big|_{\hat{\theta}} = \frac{\partial p_i}{\partial \mu_2}\Big|_{\hat{\theta}} = 0.$$

$$\frac{\partial p_i}{\partial \sigma}\Big|_{\hat{\theta}} = \hat{\sigma}^{-1}(c_{i-1}^2 e^{-\frac{1}{2}c_{i-1}^2} - c_i^2 e^{-\frac{1}{2}c_i^2})$$

$$= v_i/\hat{\sigma}.$$

Hence

$$B'B\big|_{\hat{\theta}} = \frac{M}{\hat{\sigma}^2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sum_1^M v_i^2 \end{pmatrix}.$$

The Fisher information matrix for the circular bivariate normal family is also diagonal,

$$J(\theta) = \frac{1}{\sigma^2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix},$$

so that $(J-B'B)^{-1}$ is trivially obtained. Moreover, from (3.9) it follows that

$$V_n'B = n^{-1/2}(0,0,\textstyle\sum_1^M N_i v_i/\hat{\sigma}).$$

The Rao-Robson statistic is therefore

$$R_n = X^2(\hat{\theta}_n) + (V_n'B)(J-B'B)^{-1}(V_n'B)'$$

$$= \frac{M}{n}\sum_{i=1}^{M}(N_i-\frac{n}{M})^2 + \frac{M^2}{n}\frac{(\sum_1^M N_i d_i)^2}{1-M\sum_1^M d_i^2}$$

where

$$d_i = v_i/2 = (1-\tfrac{i}{M})\log(1-\tfrac{i}{M}) - (1-\tfrac{i-1}{M})\log(1-\tfrac{i-1}{M}).$$

The limiting null distribution is $\chi^2(M-1)$, while that of the Pearson statistic $X^2(\hat{\theta}_n)$ has critical points falling between those of $\chi^2(M-1)$ and $\chi^2(M-4)$. The Rao-Robson correction term will often be necessary for a clear picture of the fit of this three-parameter family.

EXAMPLE 2. The negative exponential distribution with density function

$$f(x|\theta) = \theta^{-1}e^{-x/\theta} \qquad 0 < x < \infty$$

$$\Omega = \{\theta: 0 < \theta < \infty\}$$

is often assumed in life testing situations. Such studies often involve not a full sample, but rather Type II censored data. That is, order statistics are observed up to the sample α-quantile,

$$X_{(1)} < X_{(2)} < \ldots < X_{([n\alpha])} ,$$

where $[n\alpha]$ is the greatest integer in $n\alpha$ and $0 < \alpha < 1$. It is natural to make use of random cells with sample quantiles $\xi_i = X_{([n\delta_i])}$ as cell boundaries. Here $\xi_0 = 0$, $\xi_M = \infty$ and

$$0 = \delta_0 < \delta_1 < \ldots < \delta_{M-1} = \alpha < \delta_M = 1 ,$$

so that the $n - [n\alpha]$ unobserved X_i fall in the rightmost cell. Although the cell frequencies N_i are now fixed, the general theory of Moore and Spruill (1975) applies to this choice of cells. The use of order statistics as cell boundaries was considered by Witting (1959) and Bofinger (1973), but this application to censored data seems new. For references to previous literature on tests of fit for censored data, see Lurie, Hartley, and Stroud (1974). This example can be taken as a response to their claim that "the chi-square criterion is not generally applicable to testing the fit of Type II censored samples."

The Pearson-Fisher Statistic. Estimate θ by the grouped data MLE found as the solution of (3.2). That equation becomes in this case

$$\sum_{i=1}^{M} N_i \frac{\xi_{i-1}e^{-\xi_{i-1}/\theta} - \xi_i e^{-\xi_i/\theta}}{e^{-\xi_{i-1}/\theta} - e^{-\xi_i/\theta}} = 0$$

which is easily solved iteratively to obtain $\bar{\theta}_n = \bar{\theta}_n(\xi_1,\ldots,\xi_{M-1})$. The test statistic is

$$X^2(\bar{\theta}_n) = \sum_{i=1}^{M} \frac{[N_i - np_i(\bar{\theta}_n)]^2}{np_i(\bar{\theta}_n)}$$

where

$$N_i = [n\delta_i] - [n\delta_{i-1}] \qquad \text{(nonrandom)}$$

$$p_i(\theta) = e^{-\xi_{i-1}/\theta} - e^{-\xi_i/\theta} \qquad \text{(random)}.$$

The limiting null distribution is $\chi^2(M-2)$.

The Wald's Method Statistic. A more powerful chi-square test can be obtained by use of the raw data MLE of θ from the censored sample, namely (Epstein and Sobel, 1953),

$$\tilde{\theta}_n = \frac{1}{[n\alpha]}\left(\sum_{i=1}^{[n\alpha]} X_{(i)} + (n-[n\alpha])X_{([n\alpha])}\right) .$$

By obtaining the limiting distribution of $V_n(\tilde{\theta}_n)$ and then finding the appropriate quadratic form, a generalization of the Rao-Robson statistic to censored samples can be obtained. This is done in Mihalko and Moore (1977). The resulting statistic for the present example is

$$R_n = X^2(\tilde{\theta}_n) + (nD)^{-1}\left(\sum_{i=1}^{M} N_i v_i / p_i(\tilde{\theta}_n)\right)^2$$

where N_i and $p_i(\theta)$ are as above, and

$$v_i = \tilde{\theta}_n^{-1}\left(\xi_{i-1}e^{-\xi_{i-1}/\tilde{\theta}_n} - \xi_i e^{-\xi_i/\tilde{\theta}_n}\right)$$

$$D = 1 - e^{-\xi_{M-1}/\tilde{\theta}_n} - \sum_{i=1}^{M} v_i^2/p_i(\tilde{\theta}_n) \quad .$$

In the full sample case, $\alpha = 1$, $\xi_{M-1} = \infty$, $N_M = 0$, $\tilde{\theta}_n = \bar{X}$ and the statistic R_n reduces to the Rao-Robson statistic of Example 1, Section 3.3.3 (with M-1 cells bounded by the ξ_i).

The motivation for using censored data when lifetimes or survival times are being measured is apparent from the EXP data set. The sample 80th percentile is 9.46, while the maximum of the 100 observations is 39.12. The MLE of θ from the data censored at $\alpha = 0.8$ is $\tilde{\theta}_n = 5.471$, compared with the full sample MLE, $\bar{X} = 5.415$. Experience shows that the Roscoe-Byars guidelines are <u>not</u> adequate to ensure accurate critical points from the χ^2 distribution in the present situation, where the np_i are random and unequal. Tests of the EXP data will therefore be made with (a) the full sample using 10 cells having the sample deciles as boundaries; and (b) the data censored at $\alpha = 0.8$ using 9 cells with the first 8 sample deciles as boundaries. All cells except the rightmost in case (b) contain 10 observations. The results are, for the full sample,

$$R_n = 6.132 + 0.220 = 6.352$$

with a P-value of 0.704 from $\chi^2(9)$. For the censored sample,

$$R_n = 5.153 + 0.065 = 5.218$$

with a P-value of 0.734 from $\chi^2(8)$. These results are comparable to those obtained for the same data in Example 1 of Section 3.3.3.

REFERENCES

Bofinger, E. (1973). Goodness-of-fit using sample quantiles. *J. Roy Statist. Soc. Ser.* B 35, 277-84.

Chernoff, H. and Lehmann, E. L. (1954). The use of maximum-likelihood estimates in χ^2 test for goodness of fit. *Ann. Math. Statist.* 25, 579-86.

Cochran, W. G. (1954). Some methods of strengthening the common χ^2 tests. *Biometrics* 10, 417-51.

Dahiya, R. C. and Gurland, J. (1972). Pearson chi-square test of fit with random intervals. *Biometrika* 59, 147-53.

Dahiya, R. C. and Gurland, J. (1973). How many classes in the Pearson chi-square test? *J. Amer. Statist. Assoc.* 68, 707-12

Davis, A. W. (1977). A differential equation approach to linear combinations of independent chi-squares *J. Amer. Statist. Assoc.* 72, 212-4.

Dzhaparidze, K. O. and Nikulin, M. S. (1974). On a modification of the standard statistics of Pearson. *Theor. Probability Appl.* 19, 851-3.

Easterling, R. G. (1976). Goodness of fit and parameter estimation. *Technometrics* 18, 1-9.

Epstein, B. and Sobel, M. (1953). Life testing. *J. Amer. Statist. Assoc.* 48, 486-502.

Good, I. J., Gover, T. N. and Mitchell, G. L. (1970). Exact distributions for χ^2 and for the likelihood-ratio statistic for the equiprobable multinomial distribution. *J. Amer. Statist. Assoc.* 65, 267-83.

Lurie, D., Hartley, H. O. and Stroud, M. R. (1974). A goodness of fit test for censored data. *Comm. Statist.* 3, 745-53.

Mann, H. B. and Wald, A. (1942). On the choice of the number of class intervals in the application of the chi-square test. *Ann. Math. Statist.* 13, 306-17.

Mihalko, D. and Moore, D. S. (1977). Chi-square tests of fit for type II censored samples. Purdue University Department of Statistics Mimeograph Series No. 505.

Moore, D. S. (1971). A chi-square statistic with random cell boundaries. *Ann. Math. Statist.* 42, 147-56.

Moore, D. S. (1977). Generalized inverses, Wald's method and the construction of chi-squared tests of fit. *J. Amer. Statist. Assoc.* 72, 131-7.

Moore, D. S. and Spruill, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Ann. Statist.* 3, 599-616.

Morris, C. (1975). Central limit theorems for multinomial sums. *Ann. Statist.* 3, 165-88.

Neyman, J. (1949). Contribution to the theory of the χ^2 test. *Proc. Berkeley Symp. Math. Statist. and Prob.*, 239-73.

O'Reilly, F. J. and Quesenberry, C. P. (1973). The conditional probability integral transformation and applications to obtain composite chi-square goodness-of-fit tests. *Ann. Statist.* 1, 74-83.

Rao, K. C. and Robson, D. S. (1974). A chi-square statistic for goodness-of-fit within the exponential family. *Comm. Statist.* 3, 1139-53.

Roscoe, J. T. and Byars, J. A. (1971). An investigation of the restraints with respect to sample size commonly imposed on the use of the chi-square statistic. *J. Amer. Statist. Assoc.* 66, 755-59.

Roy, A. R. (1956). On χ^2 statistics with variable intervals. Technical Report No. 1, Stanford Univ. Department of Statistics.

Schorr, B. (1974). On the choice of the class intervals in the application of the chi-square test. *Math. Operations Forsch. u. Statist.* 5, 357-377.

Shapiro, S. S., Wilk, M. B. and Chen, H. J. (1968). A comparative study of various tests for normality. *J. Amer. Statist. Assoc.* 63, 1343-72.

Spruill, M. C. (1976). A comparison of chi-square goodness-of-fit tests based on approximate Bahadur slope. *Ann. Statist.* 4, 409-12.

Watson, G. S. (1957). The chi-squared goodness-of-fit test for normal distributions. *Biometrika* 44, 336-48.

Watson, G. S. (1958). On chi-square goodness-of-fit tests for continuous distributions. *J. Roy. Statist. Soc., Ser. B* 20, 44-61.

Watson, G. S. (1959). Some recent results in chi-square goodness-of-fit tests. *Biometrics* 15, 440-68.

Witting, H. (1959). Uber einen χ^2-Test, dessen Klassen durch geordnete Stichprobenfunktionen festgelegt werden. *Ark. Mat.* 10, 468-79.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer